LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-412375

# Molecular Dynamics Simulations of Highly Charged Green Fluorescent Proteins

E. Y. Lau, J. L. Phillips, M. E. Colvin

April 21, 2009

Molecular Physics

## Disclaimer

# Molecular Dynamics Simulations of Highly Charged Green Fluorescent Proteins

Edmond Y. Lau[1], Joshua L. Phillips[2], and Michael E. Colvin[3*]

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550

[2]Center for Computational Biology and School of Natural Sciences, University of California, Merced, Merced, California 95344

[3]Center for Computational Biology and School of Natural Sciences, University of California, Merced, Merced, California 95344

Corresponding author: mcolvin@ucmerced.edu

Abstract

A recent experimental study showed that green fluorescent protein (GFP) that has been mutated to have ultra-high positive or negative net charges, retain their native structure and fluorescent properties while gaining resistance to aggregation under denaturing conditions. These proteins also provide an ideal test case for studying the effects of surface charge on protein structure and dynamics. We have performed classical molecular dynamics (MD) simulations on the near-neutral wildtype GFP and mutants with net charges of -29 and +35. We analyzed the resulting trajectories to quantify differences in structure and dynamics between the three GFPs. This analysis shows that all three proteins are stable over the MD trajectory, with the near-neutral wild type GFP exhibiting somewhat more flexibility than the positive or negative GFP mutants, as measured by the order parameter and changes in phi-psi angles. There are more dramatic differences in the properties of the water and counter ions surrounding the proteins. The water diffusion constant near the protein surface is closer to the value for bulk water in the positively charged GFP than in the other two proteins. Additionally, the positively charged GFP shows a much greater clustering of the counter ions (Cl-) near its surface than corresponding counter ions (Na+) near the negatively charged mutant.

## Introduction

A recent paper by Liu and coworkers reports the interesting result that adding significant positive or negative charges to the surface of a globular protein does not significantly affect the protein's function, structure or stability, but does impart significant resistance to aggregation upon chemical or thermal denaturing. Specifically, Liu mutated neutral surface amino acids on the 248 amino acid green fluorescent protein (GFP) to either positively charged amino acids (lysine or arginine) or negatively charged amino acids (aspartic or glutamic acid) to yield mutant GFPs with net charges ranging from -30 to +48, compared to a standard GFP charge of -7. All GFPs, regardless of charge, where found to have similar fluorescence and circular dichroism spectra, indicating minimal structural differences. Based on the guanidinium-induced denaturation curves, Liu reports similar unfolding free energies of 11.2, 10.2, and 8.8 kcal/mole for the GFPs with net charges of -7, -30 and +36, respectively. Such highly charged mutant proteins may have practical value in guiding the engineering of aggregation-resistant proteins for biotechnological work. Scientifically, these proteins provide an excellent model system for probing the effects of surface charge on protein structure and protein-solvent interactions. Moreover, since these proteins have net charges far outside the usual range of -10 to +10 seen for wildtype proteins, they may provide new insights on the biological constraints on protein net charge in living organisms. The goal of this study is to use classical molecular dynamics to evaluate the effects of the very high surface charges on the structure, dynamics, and solvation shell for GFPs with a wide range of surface charges.

Methods

We started with the 1.9 Å resolution X-ray crystal structure of GFP published by Yang, et al. which is available in the Protein Data Bank (http://www.rcsb.org) with PDB ID 1GFL. The protein in this crystal structure differs from the reference structure used by Liu in their mutation studies at 6 point mutations that convert this to the "superfolder" variant. Additionally, the 1GFL sequence does not have the 10 amino acid histidine tag at the N-terminal end and 9 residues at the C-terminal end. The program Modeller was used to create highly charged mutants of superfolder GFP based on this structure. We modeled the mutations necessary to form the superfolder GFP for our unmutated GFP and those corresponding to the -30 and +36 GFPs synthesized by Liu. The aligned sequences for all 3 mutants and the original 1GFL are given in Supplementary Figure 1. In this paper we will refer to the unmutated GFP as "sGFP", the highly negatively charged mutant as "nGFP" and the highly positively charged mutant as "pGFP". Figure 1 shows solvent accessible surfaces for all three GFPs with charged residues color coded as follows: Red--negative; Blue--positive; Grey--Uncharged. Note that because of the missing amino acids at the C-terminal end, our GFPs have slightly different net charges than Liu's. Our sGFP has a net charge of -6 (versus -7 in Liu's study), nGFP has a charge of -29 (versus -30), and pGFP has a charge of +35 (versus +36). Additionally, the simulated structures did not include the fluorophore which is formed by a post-translational reaction of three residues (Ser-Tyr-Gly). Hence, our simulated structures correspond to the protein product immediately after translation and folding.

Using the LEAP program distributed with Amber 8.0, each GFP structure was placed in a rectangular periodic box with a minimum spacing of 10 Å between the protein

and the box edge. Each protein was neutralized with just the necessary number of Na+ or Cl- counter ions using the default procedure in the LEAP module that calculates the electrostatic potential on a 1 Å grid around the solute and then places the ions at the points of lowest (or highest) electrostatic potential. For the highly charged GFP mutants, this leads to the ions being placed in close proximity to the protein surface, for example, the initial $Cl^-$ placement around pGFP is shown in Supplementary Figure 2. The box was then filled with TIP3P waters. Additional ions were not added to the box to avoid the noted nucleation problem with NaCl in solution using Amber. The resulting total system sizes were 25,000-28,000 atoms and approximately 78 x 69 x 61 $Å^3$.

Minimization and molecular dynamics were performed using the ff99 force field. and the pmemd module of the Amber 8.0 program suite. First, 500 steps of minimization were performed using steepest descent and conjugate gradient. Each molecular dynamics simulation was run to 10 ns using an NPT ensemble with Berendsen's weak temperature coupling method (2ps time constant) and isotropic volume scaling (1 ps time constant). A time step of 2 fs was used and SHAKE was applied to all bonds involving hydrogen. The electrostatic interactions were treated by particle mesh Ewald methods using an 1 Å grid and 9 Å cutoff for van der Waals interactions. Structures were saved every half picosecond.

The high charge on these proteins means that the equilibration time of the counter ions may be more similar to that needed for DNA (which carries a -2 charge for every base pair) than for conventional proteins. Previous studies on the dynamics of counter ions around DNA have shown that multiple nanoseconds are needed for their equilibration. For example, Ponomarev, et. al found that the DNA-sodium radial

distribution function required up to 5 ns to fully stabilize. An earlier study by Feig and Pettitt also concluded that "at least several nanoseconds" were required for ion equilibration. For this reason, we allowed the first 5 ns of our simulations for equilibration, and all analysis was performed on the final 5 ns.

A 10 ns NaCl simulation was performed to obtain comparison data. Five Na+ and Cl- ions were solvated in 1945 TIP3P waters (40 x 40 x 40 $Å^3$) to give an approximately 0.15 M solution. The simulation was performed using the protocols outlined above.

The stability and structural similarity of the GFPs were determined using several methods for comparing protein structure. The first method we used was the standard root mean squared distance (RMSD) between the $C_\alpha$ atoms on the protein backbone. We first computed the quaternion (and corresponding left-rotation matrix) that minimizes the Euclidean distance between the corresponding pairs of $C_\alpha$ atoms, then rotated the structure using this transformation to yield the optimal superposition, and finally calculated the RMSD between corresponding $C_\alpha$ pairs.

To determine if there was any evidence for change in fold over the simulation, we used MAMMOTH, a tool developed for determining the fold similarity of two protein structures. MAMMOTH works by creating multiple superpositions of the two protein structures to determine a statistical measure, in the form of a Z-score, indicating whether the two folds are the same. We calculated the MAMMOTH score for structures sampled every 5ps from the last 5ns of the trajectories, using the initial structure as the reference.

We calculated ensemble-averaged order parameters ($S^2$) for all bonds between the amide nitrogen and hydrogen along the backbone of the protein using the following formula.

$$S_{ij}^2 = \frac{3}{2}\left( \left\langle \hat{x}_{ij}{}^2 \right\rangle^2 + \left\langle \hat{y}_{ij}{}^2 \right\rangle^2 + \left\langle \hat{z}_{ij}{}^2 \right\rangle^2 + 2\left\langle \hat{x}_{ij}\hat{y}_{ij} \right\rangle^2 + 2\left\langle \hat{x}_{ij}\hat{z}_{ij} \right\rangle^2 + 2\left\langle \hat{y}_{ij}\hat{z}_{ij} \right\rangle^2 \right) - \frac{1}{2}$$

where $S_{ij}^2$ is the order parameter for the bond from atom *i* to atom *j* while $x_{ij}$, $y_{ij}$, and $z_{ij}$ denote the components of the vector along this bond. The order parameter indicates how broad the distribution of vectors is across the ensemble. A value of zero means that the vectors are uniformly distributed while a value of one indicates that the distribution is degenerate.

As a measure of the rate of overall structural change, we calculated the autocorrelation function of the vector of backbone φ-ψ angles. We first create a high dimensional vector encoding all of the φ-ψ angles by first transforming the angle vectors into a form that correctly preserves the similarity or rotational symmetry of vector angles near the edges of the range [-π,π]. This is accomplished by building a vector of complex numbers whose angular components (arguments) are directly taken from the vector of φ-ψ angles and which has a sequence symmetry commonly referred to as *conjugate-complex*. The modulus of all components of this complex vector are set to one, and an additional value of 1+0i is pre-pended to the vector. Applying the inverse Discrete Fourier Transform to this conjugate-complex vector results in a unit-length vector of real numbers. The dot product can be used to compute a measure of similarity between two of these vectors which takes into account the rotational symmetry of the backbone angles. The autocorrelation function was calculated over a time window of 1000 ps and averaged over all time origins.

The mean squared displacements (MSD) for bulk solvent were calculated in 500 ps windows over the last 5 ns of dynamics and averaged. Diffusion constants for the bulk solvent were estimated from the slope of the MSD. Hydration (close) water was defined

as by having the oxygen of the water molecule within 3.4 Å of any of the peptide backbone heavy atoms (C, $C_\alpha$, N). The diffusion constants for the hydration waters were estimated with 500 ps windows over the last 5 ns and averaged. Hydration pattern around the proteins were determined using the grid option in ptraj over the last 5 ns of dynamics. A cutoff of 385 was used for the water number density and 10 for both the sodium and chloride number density. The solvent accessible surface area was calculated using the program DSSP using a probe radius of 1.4 Å.

The effect of the high charges in GFP on the water was studied by determining an angle, θ, between two vectors, one vector is formed from the α-carbon to the nearest water oxygen and the second vector is formed from the water oxygen to the midpoint between the hydrogens of the water (i.e. the dipole moment vector for the water). The analysis of the water orientation angles was calculated only out to 20 Å from the α-carbons on the surface of the protein to minimize the effects of periodic images on the water orientation.

## Results and Discussion

### GFP Structure and Dynamics

All three proteins were stable during the 10ns simulations and showed similar RMSD changes over time. Figure 2 shows the RMSD values, plotted versus time for the final 5 ns of each trajectory, using the structure at the beginning of the final 5 ns as the reference. Figure 2a shows the RMSD for the full length (230 residue) GFPs and indicates a larger and steadily growing amount of structural change for pGFP as compared to sGFP and nGFP. Visualization of the trajectories showed that this large

variation was limited to the terminal residues. Figure 2b shows the RMSD plots based on the central 222 residues, excluding the four terminal residues at each end of the proteins. This plot shows all three GFPs having asymptotic RMSDs of about 1 Å, indicating that all three are structurally stable over the simulations and, further, show a similar rate of average backbone change over the 5ns of simulation. Earlier MD studies of the wild type GFP and single residue mutants have found a very similar asymptotic value for the $C_\alpha$ RMSD of < 1 Å using different force fields. The RMSD plots for the full 10ns of simulation are shown in Supplementary Figure 3. These show very similar trends to the RMSD plot for the final 5 ns, but the asymptotic RMSD is increased by 0.5 Å.

To evaluate the magnitude of the structural change during the entire simulation process (mutation, minimization and dynamics), we calculated the average RMSD between the crystal structure and structures sampled every 50 ps from the final 5 ns seconds of each trajectory. These results are shown in the upper portion of Table 1. For the entire 230 amino acid sequence, the highly charged mutant GFPs (nGFP and pGFP) exhibit about 25% increased RMSD from the crystal structure relative to sGFP. In all cases the $C_\alpha$ RMSD are reasonably small (<2.5Å), indicating no dramatic change in the backbone configuration due to the mutation process. Additionally, the different structures were compared using the average RMSD between all pairs of structures sampled every 50 ps. These results are shown in the lower portion of Table 1. This shows that the each structure is about equally different from each other, 1.6-1.8Å $C_\alpha$ RMSD, and have an overall similar backbone conformation. Table 1 also shows the $C_\alpha$ RMSD results calculated for only the central 222 amino acids. These results are very

similar to those calculated for the entire sequence, with little variation (<1.64 Å) from the crystal structure and little variation (<1.64 Å) between the different GFP mutants studied.

The results from MAMMOTH for the change in fold show some variation between the GFPs. For nGFP, there was no measurable difference in the fold over the entire 5ns. For sGFP there were only 9 of 1000 sampled structures that showed a measurable differences and these differences were very small, (less than ½% change in Z-score). For the pGFP, more than half the sampled structures showed a measureable change in Z-score, although the highest changes were less than 2%. Overall, no significant fold changes were observed during the simulations.

We evaluated the amount of structural fluctuation in the different GFPs using several techniques. The root mean squared fluctuations (RMSF) of the backbone atoms of the different GFPs were very similar between the simulations and in good agreement with the fluctuations estimated from the experimental crystallographic B-factors (Figure 3) for the wildtype GFP. The termini were very mobile in the simulations because in the starting structure they are extended coils. The N-terminus of pGFP had the largest RMSF but this was not due to large scale motion. Instead, the N-terminus of pGFP folded into a 2-turn alpha helix which caused large scale changes in the position of the terminal residues. As the plot show, all of the major B factor peaks seen in the experimental data are also predicted in the MD simulations for all three GFPs. An earlier MD study using the CHARMM22 force field of wildtype GFP and a single residue mutant found a very similar level of agreement with the experimental B factors. The relative magnitude of the fluctuations for each GFP varies, but in nearly all the peaks, the predicted fluctuations in

sGFP are at least as large as those for nGFP or pGFP. This shows that there is no systematic increase in the surface fluctuations due to the increased surface charge.

The amide bond (N-H) order parameter is another experimentally accessible measure of the dynamical order of a protein structure. Figure 4 shows plots of the order parameter versus residue for each of the three GFPs. Note that since $S^2$ is essentially the time autocorrelation for the bond vector, a low value of $S^2$ indicates a high level of disorder. The plot in Figure 4 shows that the wild type sGFP is predicted to have higher fluctuations at nearly every residue than the highly charged nGFP or pGFP mutants. Additionally, at most residues, the positively charged pGFP is predicted to show higher fluctuations than nGFP.

The results for our third measure of protein dynamics, the φ-ψ vector autocorrelation function, are shown in Figure 5. These autocorrelation functions show that protein dynamics is occurring on two time scales. On a very short time scale (~10ps) are thermal fluctuations that lead to a ~5% drop in the autocorrelation function. On a much longer time scale there is a slow, approximately linear decrease in that corresponds to a steady rotation of the φ-ψ vector. For the rapid time scale, the wild type sGFP has the largest drop in the autocorrelation function indicating higher thermal fluctuations compared to the two highly charged mutants. For the slower rate of slow change, all three GFPs show a very nearly linear decay in the autocorrelation function. ($R^2$ for a linear fit is 0.99 for $\Delta t$=250-1000 ps.) The slope of this long-time decay is -1.15, -1.08, and -1.05 for sGFP, nGFP, and pGFP, respectively (in units of $10^{-5}$ $ps^{-1}$). Consistent with the previous measures of structural fluctuation, this shows that the wildtype sGFP

undergoes the most rapid structural change, followed by the negatively charged nGFP and finally the positively charged pGFP.

Water and Counter Ion Distribution

It is well established that proteins perturb their immediate solvent environment, affecting both the diffusion and distribution of counter ions and water. This set of three proteins with very similar structures, but very different surface charges, provide an ideal test case for evaluating how net protein charge affects the water and counter ions near the protein surface. The water densities near the protein surface are given in the second and third columns of Table 2. Also, the number density for water near the GFP surface are shown in Figure 6. These results show that having a high negative charge around nGFP does not significantly perturb the hydration layer around the protein when compared to sGFP. The average number of water molecules within 3.4 Å of the heavy backbone atoms is 83.9 and 85.4 for sGFP and nGFP, respectively (see Table 2 and Figure 6).

The pGFP simulation does show a difference in hydration. There are on an average of 90.5 water molecules in close contact with the protein. Calculation of the water number density around each protein clearly shows the difference in hydration between the proteins. Interestingly, there is a decrease in hydration in the pGFP simulation relative to the sGFP and nGFP simulations when counting waters within 5.0 Å of the protein surface (see Table 2). Examination of the pGFP protein structure shows that the arginine and lysine sidechains are predominately in an extended conformation due to the high charge concentration and the close proximity of the sidechains with one another. For pGFP, 53 out of 230 residues are either arginine or lysine. On average, the lysine sidechain is extended 5.92 Å from the surface ($C_\alpha$ to amine N distance) and the

arginine sidechain is extended 5.71 Å from the surface ($C_\alpha$ to guanidinium C distance). The shortest average distance was for Arg96 (4.97 Å) and 15 of the Arg/Lys had average sidechains lengths greater than 6 Å. These sidechains take up the volume that would otherwise be filled with water molecules in the region between 3.4 and 5.0 Å from the surface.

The difference in net surface charge also has a strong effect on the average orientation of the bulk water. In all the GFP simulations, the water molecules in closest to the α-carbons have the hydrogens oriented towards the protein with the maximum θ of about ~60° (Figure 7A). The further away from the surface of the protein the greater the influence of the protein charge on the orientation of the water. This is especially pronounced around 5 Å from the protein surface (which corresponds to the average length of Arg and Lys sidechain) in the pGFP simulation. In the sGFP and nGFP simulations the dipole moment of waters tend to pointed away from the protein surface although sGFP shows a weaker effect. In the pGFP simulation, the opposite occurs and the dipole moment of water molecules tend to pointed towards the protein. The effect is more clearly seen with a cumulative cos(θ) which is the sum of cos(θ) of all the water molecules as a function of distance (Figure 7B). These highly charged proteins are able to affect the orientation of water through out the simulation system. Since these simulations were performed with periodic boundaries this could have enhanced the orientation preference. Much larger water boxes and a solution containing both sodium and chloride maybe necessary to further study this effect. Although the orientation preference of the waters differ in the simulations this did not significantly affect their dynamics.

A number of earlier molecular dynamics studies have shown that the interactions with the protein surface (especially polar residues) can retard water diffusion. Although the diffusion coefficient of the total water in these GFP simulation is not significantly different than obtain from a simulation of bulk water, the diffusion constant of the waters nearest the protein is decreased. The diffusion constants for water both near the protein surface (<3.4Å) and in the bulk are listed in columns 4 and 5 of Table 2. The diffusion coefficients of hydration water in the sGFP and nGFP simulations are similar at 3.78 and 3.86 ($10^{-5}$ $cm^2/s$), respectively. Interestingly, the diffusion coefficient of the hydration waters in the pGFP simulation (4.46 $10^{-5}$ $cm^2/s$) is much closer to the bulk value. This difference in diffusion coefficient for the hydration water is likely due to the hydrophobicity close to the pGFP protein backbone.

Although the overall charge of the protein is high, the charges are concentrated at the end of the Arg or Lys sidechain and the rest is aliphatic. In fact, the long sidechains of the Arg and Lys increase the solvent accessible surface area (SASA) of the protein. The average (and standard deviations) of the SASA over the last 5ns of dynamics for sGFP, nGFP, and pGFP were 11742.0 (149.8), 12093 (191.7), and 13114.3 (179.9) $Å^2$. There is an approximately 12% increase in SASA, most of which is hydrophobic, for pGFP relative to sGFP. Having the Arg and Lys sidechains extended prevents water in contact with the protein surface from interacting with the hydrophilic portion of the sidechain. This likely contributes to the higher diffusion coefficient for water close to the surface of pGFP. Polar residues and Asp/Glu are about half the sidechain length and do not exhibit a similar effect.

Figure 8 shows the distribution of counter ions in both nGFP and pGFP. The interaction of the counter ions with the supercharged GFP differ. Sodium ions did not show any preferred interaction site with nGFP. Inspection of the sodium number density show the ion to be distributed through out the system. The chloride ions in the pGFP simulation do show a preference to interact with the protein. The interaction takes place at the charged end of the Lys/Arg sidechains, and is probably a result of the extension of these sidechains away from the protein surface.

## Conclusion

In this study we have analyzed the structure, dynamics and solvent interactions of GFP and two highly charged mutants using classical Molecular Dynamics. Our results show relatively minor differences in the structure and dynamics of the three GFPs, although the near neutral wild type (sGFP) is predicted to have increased flexibility based on our measured B factors, order parameters and autocorrelation of the vector of $\varphi$-$\psi$ angles. We observed more significant differences in the protein-solvent interactions. In particular, we find the charged residues on the pGFP are fully extended away from the protein surface and into the solvent. This leads to differences in the local water density, diffusion constants, and counter ion distributions between pGFP and the other two mutants.

## Acknowledgements

This work was also supported by the U.S. Department of Energy, Office of Science, Offices of Advanced Scientific Computing Research, and Biological & Environmental

Research through the U.C. Merced Center for Computational Biology #DE-FG02-04ER25625. This work was in part performed under the auspices of the United States Department of Energy by Lawrence Livermore National Laboratory under contract number DE-AC52-07NA27344.

Tables:

Table 1: Upper portion: Average RMSD (Å) relative to the crystal structure for the three GFPs sampled every 50ps from the 5 ns MD trajectories. Lower portion: Average RMSD (Å) between different GFP trajectories sampled every 50 ps. In both portions of the table the middle column shows the RMSD calculated from all 230 amino acids and the rightmost column shows the RMSD calculated from the middle 222 amino acids.

| Comparison | Average RMSD Å (st. dev. in parentheses) | Average RMSD Å based on central 222 amino acids (st. dev. in parentheses) |
|---|---|---|
| Crystal-sGFP | 1.724 (0.163) | 1.361 (0.082) |
| Crystal-nGFP | 2.104 (0.092) | 1.639 (0.067) |
| Crystal-pGFP | 2.226 (0.345) | 1.467 (0.122) |
| sGFP-nGFP | 1.562 (0.135) | 1.362 (0.117) |
| sGFP-pGFP | 1.627 (0.236) | 1.225 (0.107) |
| nGFP-pGFP | 1.793 (0.141) | 1.428 (0.109) |

Table 2: Columns 2-3: number of waters within 3.4 Å and 5.0 Å of the protein surface (distance from backbone heavy atoms). Columns 4-5 Diffusion constants (in units of $10^{-5}$ $cm^2/s$) for water molecules close to the protein surface and in bulk. Columns 6-7: Diffusion constants (in units of $10^{-5}$ $cm^2/s$) for counter ions.

| Protein | #$H_2O$ <3.4 Å | #$H_2O$ <5.0 Å | D($H_2O$) <3.4 Å | D($H_2O$) | D($Na^+$) | D($Cl^-$) |
|---|---|---|---|---|---|---|
| sGFP | 83.9 (6.3) | 472.7 (11.1) | 3.78 | 4.93 | -- | -- |
| nGFP | 85.4 (6.7) | 481.1 (11.6) | 3.86 | 5.00 | 1.52 | -- |
| pGFP | 90.5 (6.6) | 462.2(12.0) | 4.46 | 5.05 | -- | 2.72 |
| Bulk | | | | 5.19 | 2.38 | 3.32 |

Figure Captions

Figure 1: Solvent accessible surfaces for three GFPs with charged residues color coded. Red: negative; Blue: positive; Grey: Uncharged.

Figure 2: Plots of root mean square distances (RMSD) of the backbone $C_\alpha$ atoms calculated relative to the first structure in the 5ns simulation. a) RMSD versus time for the full 230 residue GFPs; b) RMSD versus time for the central 222 residues of the GFPs.

Figure 3: Plots of the theoretical and experimental B-factors ($Å^2$) versus residue position. The experimental B factors are for the wild type GFP.

Figure 4: Plot of predicted amide bond (N-H) order parameter ($S^2$) versus residue position. Lower values of $S^2$ indicate higher levels of fluctuation.

Figure 5: Plots of the transformed φ-ψ autocorrelation function versus Δt for the three GFPs.

Figure 6: Number density of water around the three GFPs studied. The enhancement of hydration water around pGFP is approximately uniform.

Figure 7: Panel 7A shows the plot of cos(θ) for water as a function of distance from the protein surface (α-carbons). Panel 7B plots the cumulative cos(θ) for all the waters as a function of distance.

Figure 8: Plot of the number density of sodium (nGFP) and chloride (pGFP) around GFP.

Figure 9. Number density of chloride in the vicinity of the Arg and Lys sidechains in the pGFP simulation.

Supplementary Figure 1: Alignment of amino acids sequences used for all three GFPs (sGFP= "STD", nGFP= "NEG", pGFP= "POS") and the sequence of the protein whose crystal structure was used as a structural template (1GFL).

Supplementary Figure 2: Initial distribution of $Cl^-$ ions around the pGFP as placed by the LEAP module of AMBER 8.0, that calculates the electrostatic potential on a 1 Å grid around the solute and then places the ions at the points of lowest (or highest) electrostatic potential.

Supplementary Figure 3: Plots of root mean square distances (RMSD) of the backbone $C_\alpha$ atoms calculated relative to the first structure in the full 10ns simulation. a) RMSD versus time for the full 230 residue GFPs; b) RMSD versus time for the central 222 residues of the GFPs.

References

Figure 1

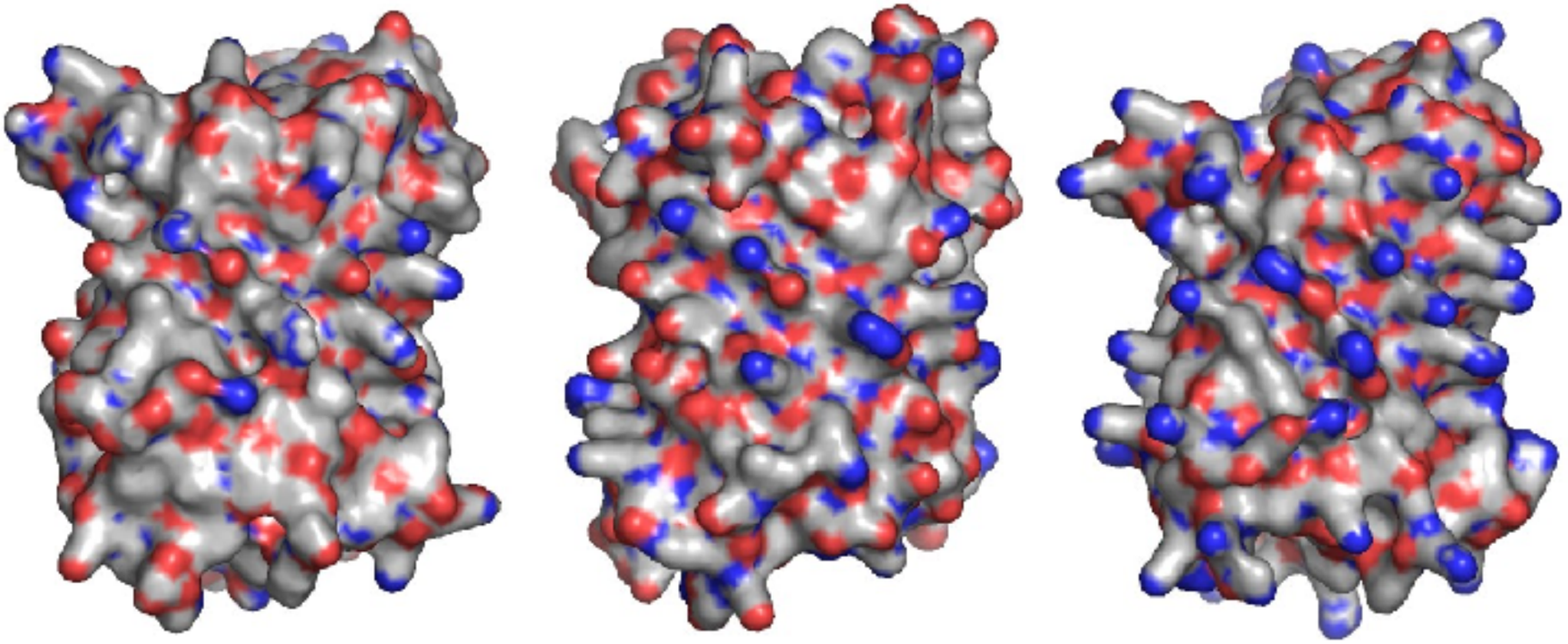
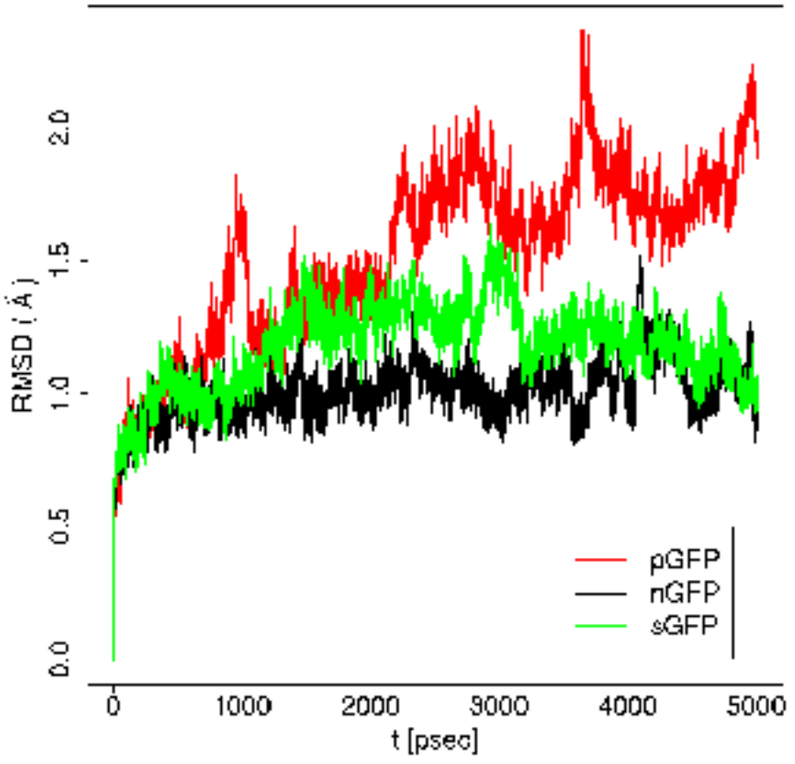


Figure 2


RMSD ( Å )
0.0
0.5
1.0
1.5
2.0
0
1000
2000
3000
4000
5000
t [psec]
pGFP
nGFP
sGFP

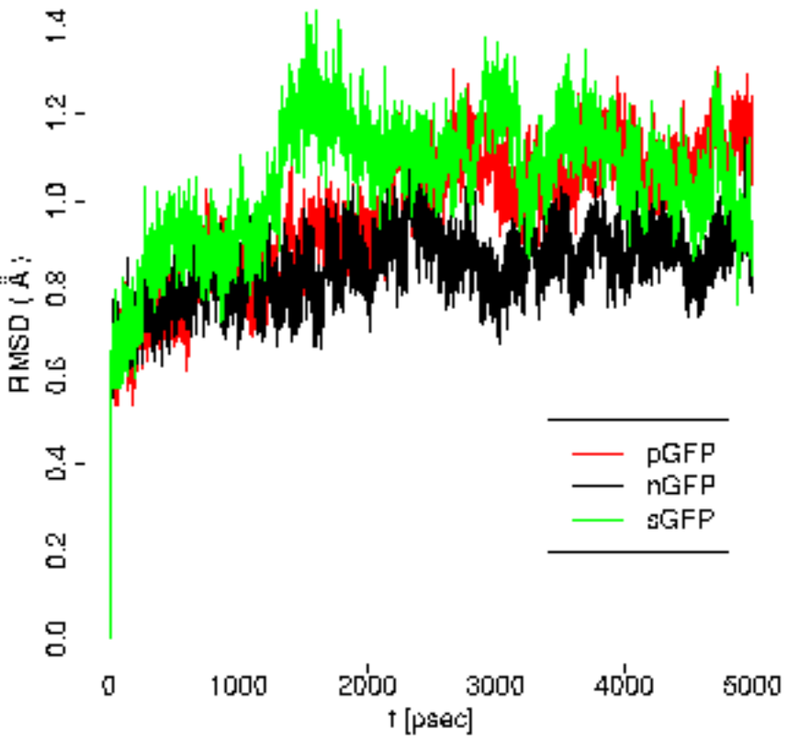

RMSD ( Å )
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
1000
2000
3000
4000
5000
t [psec]
pGFP
nGFP
sGFP

Figure 3

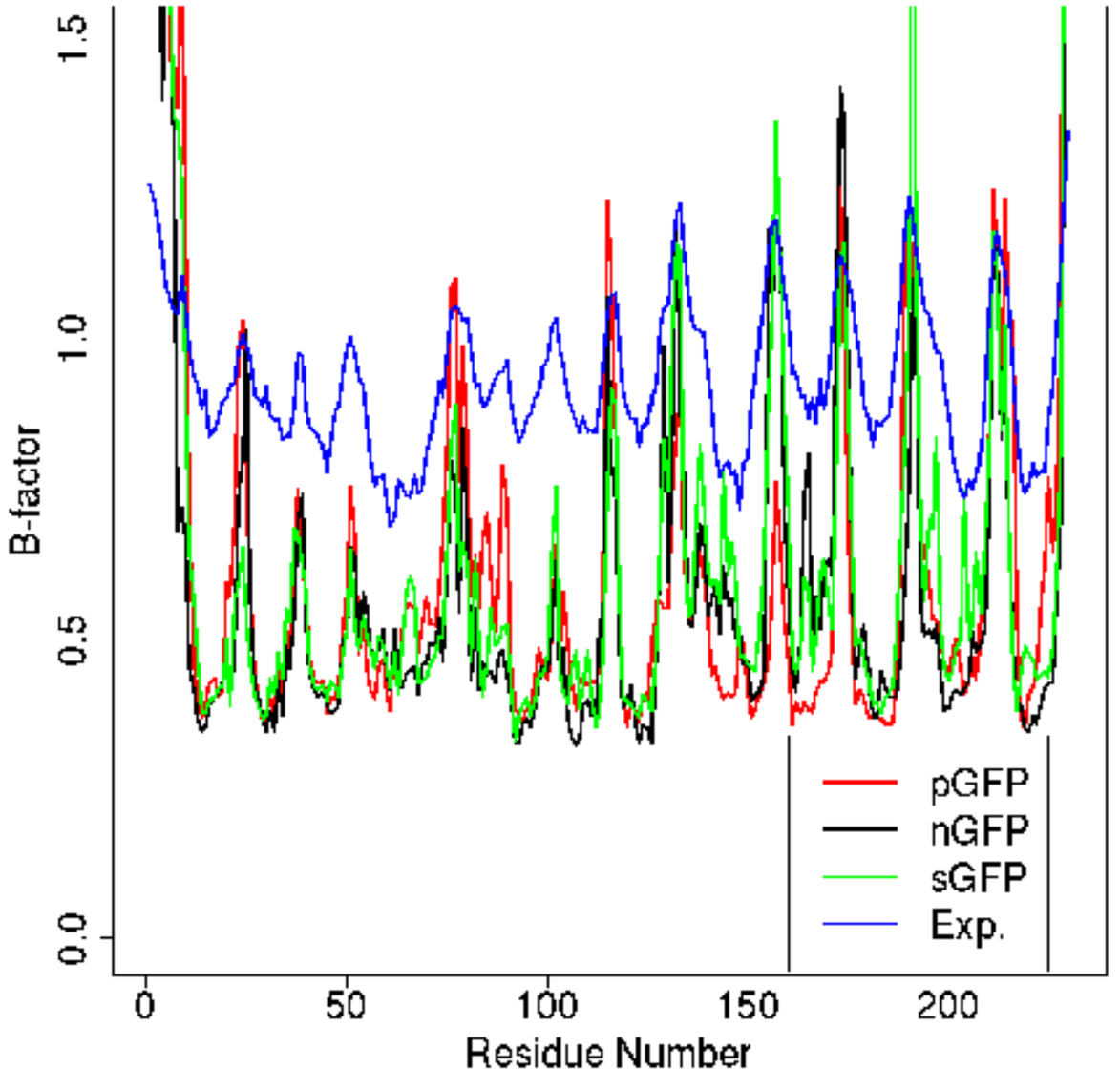

B-factor
0.0
0.5
1.0
1.5
0
50
100
150
200
Residue Number
pGFP
nGFP
sGFP
Exp.

Figure 4

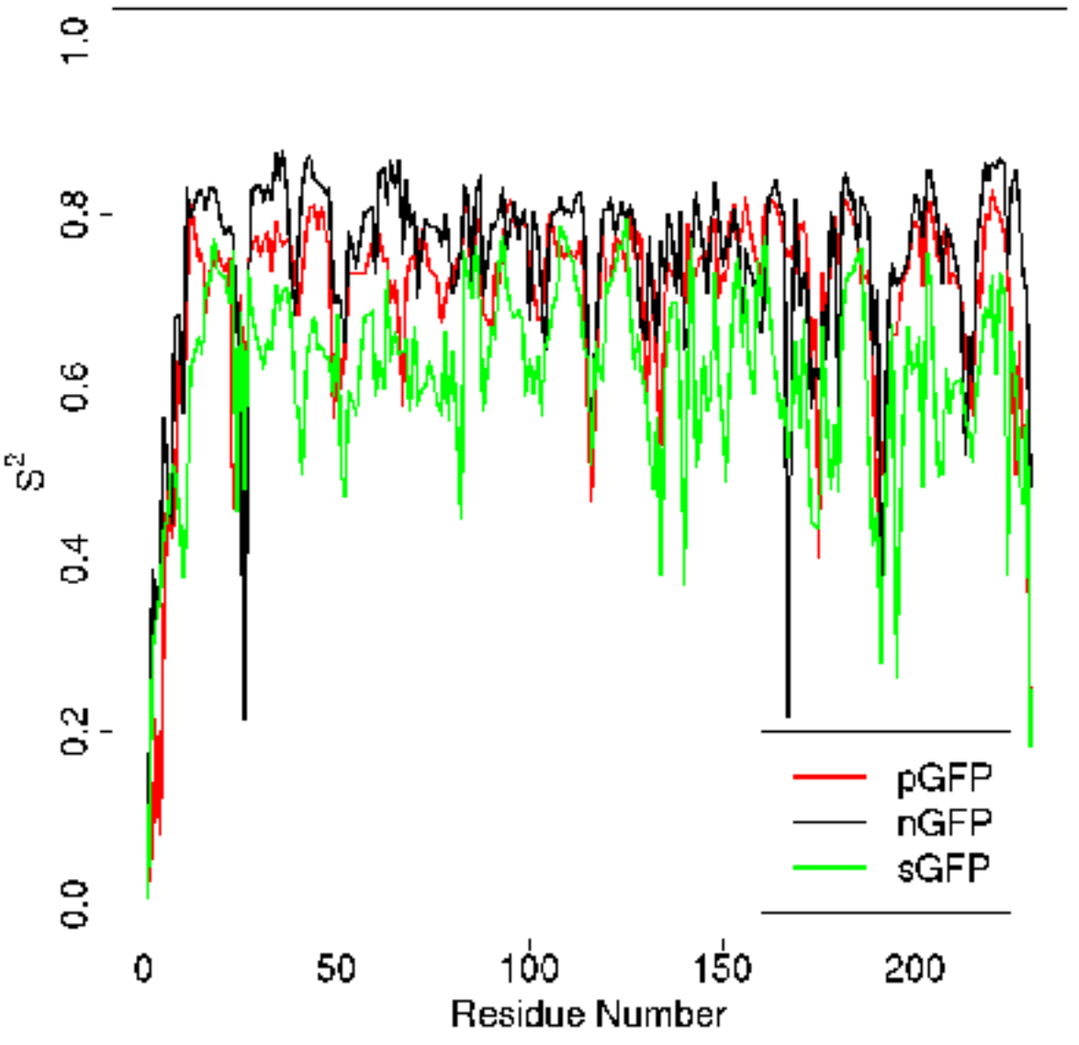

1.0
0.8
0.6
0.4
0.2
0.0
$S^2$
0
50
100
150
200
Residue Number
pGFP
nGFP
sGFP

Figure 5

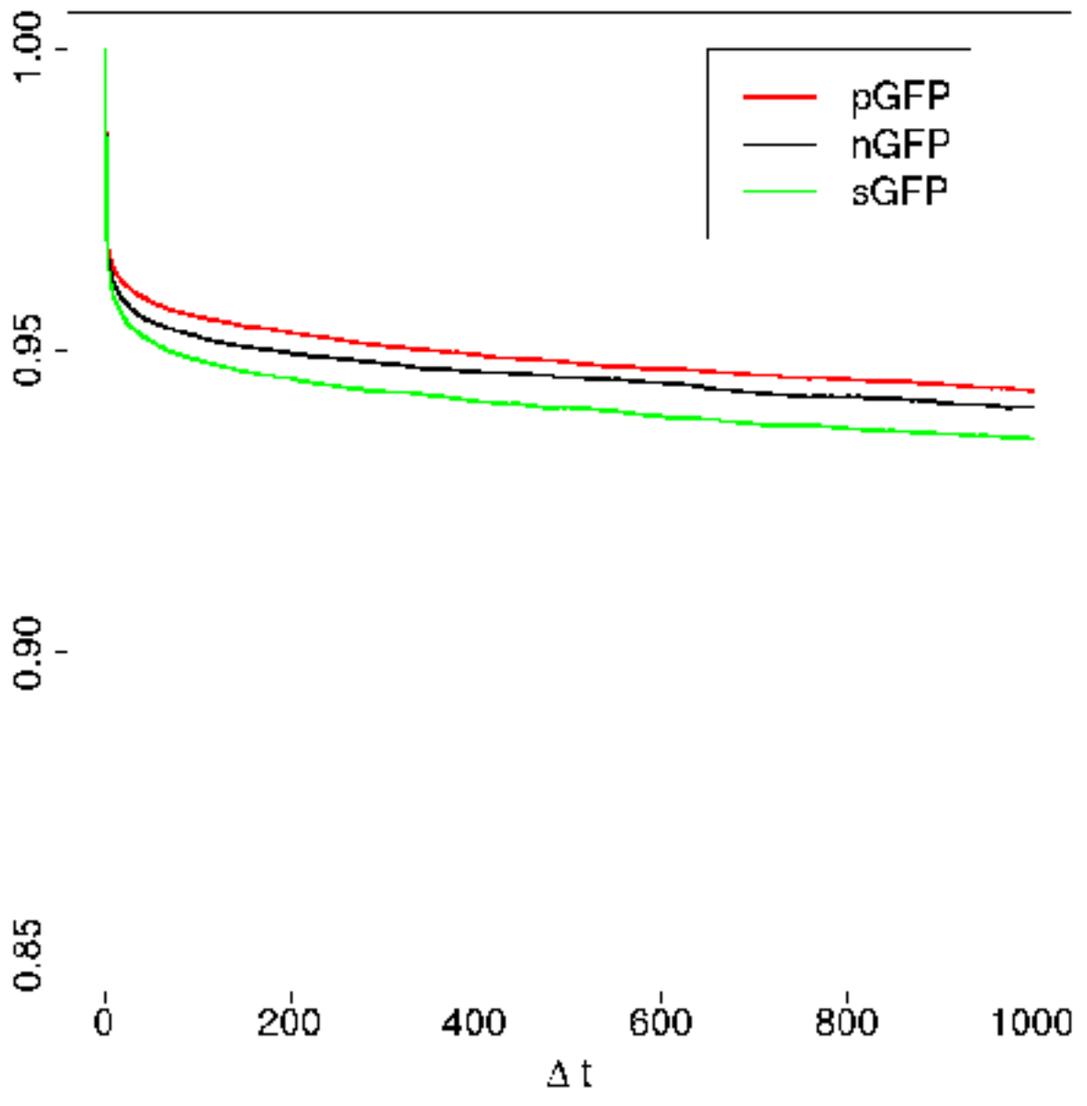
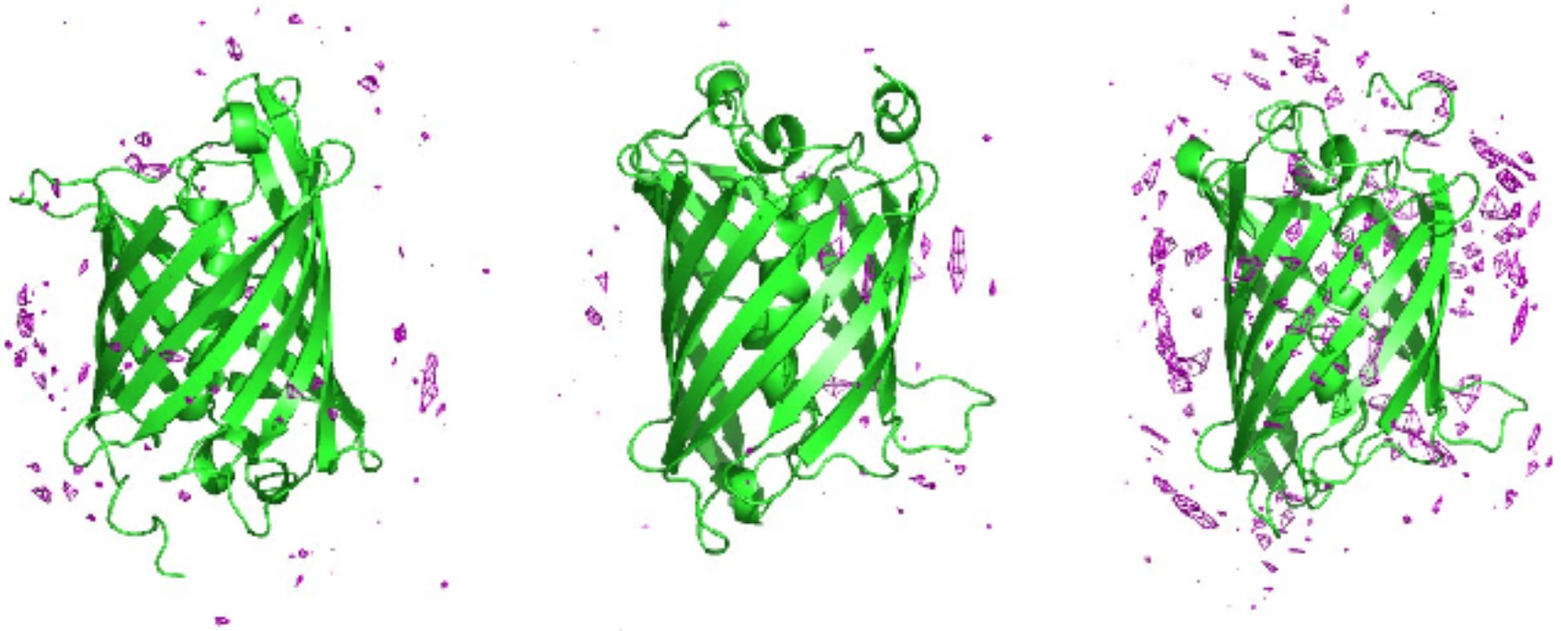


Figure 6



sGFP nGFP pGFP

Figure 7a

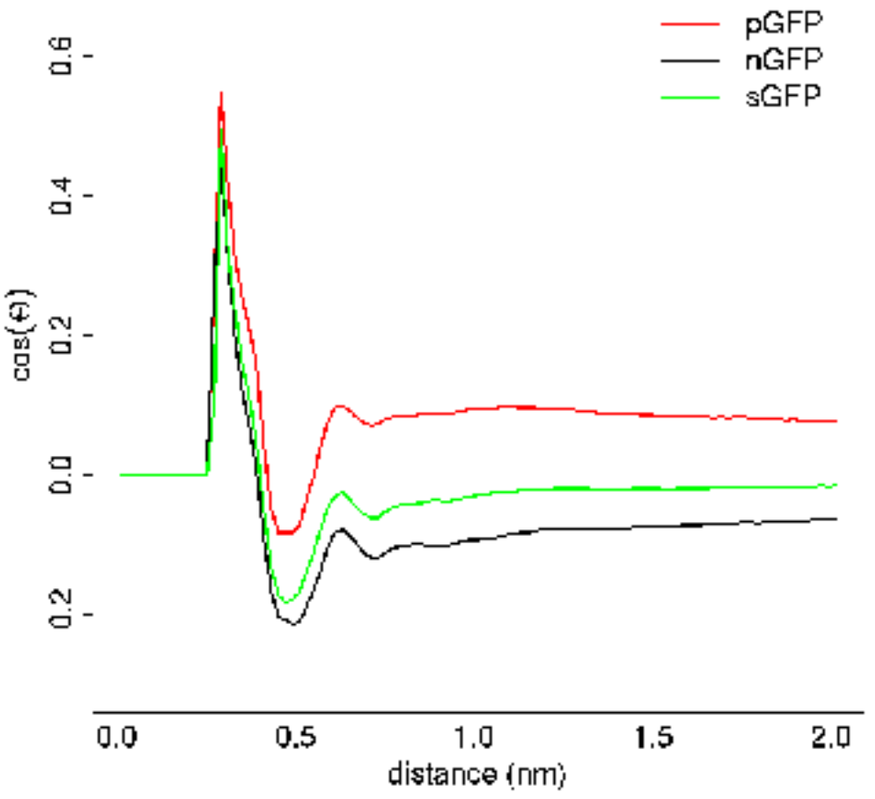


Figure 7b

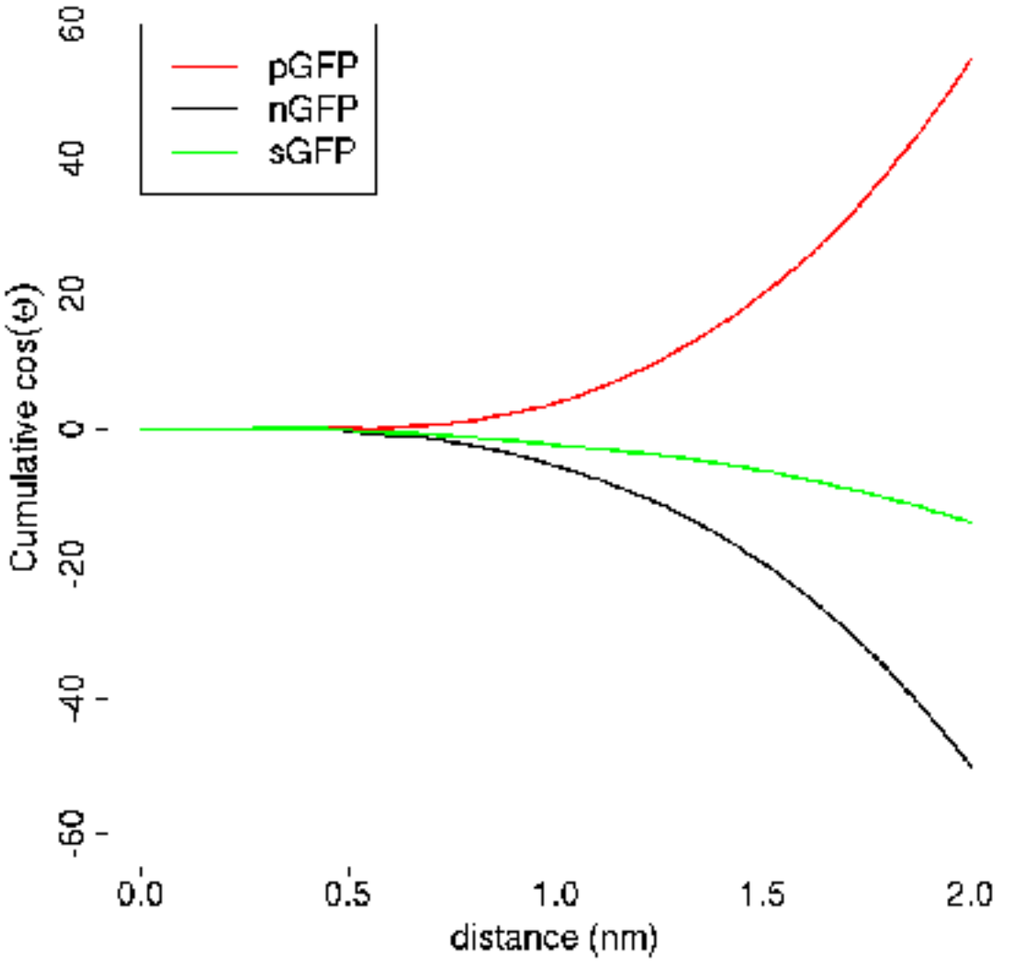
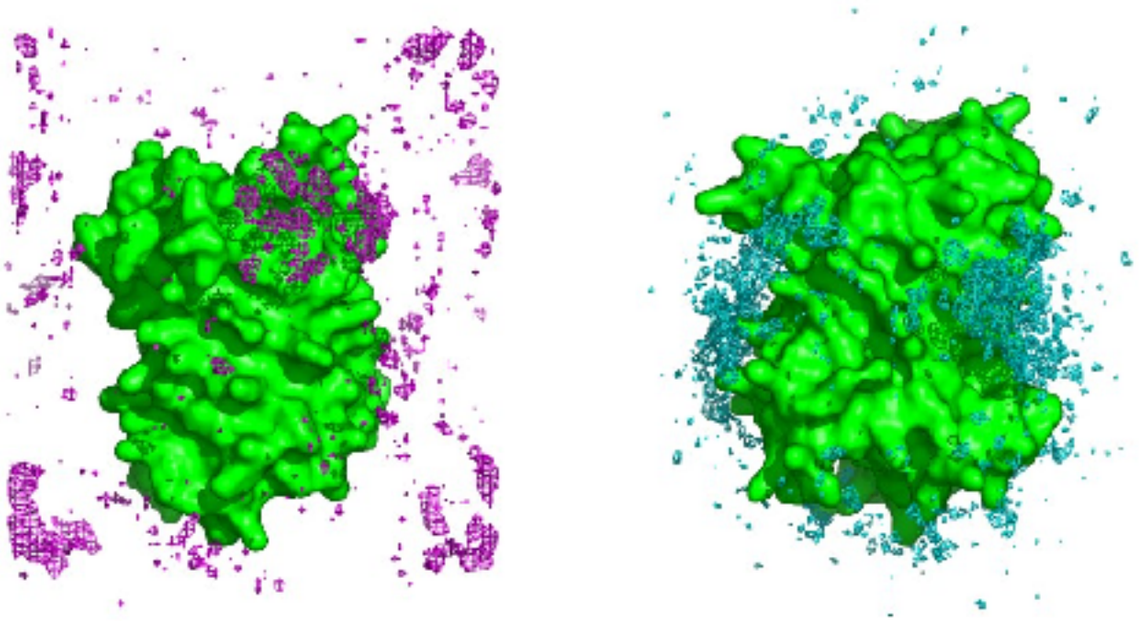

pGFP
nGFP
sGFP
Cumulative cos(θ)
60
40
20
0
-20
-40
-60
0.0
0.5
1.0
1.5
2.0
distance (nm)

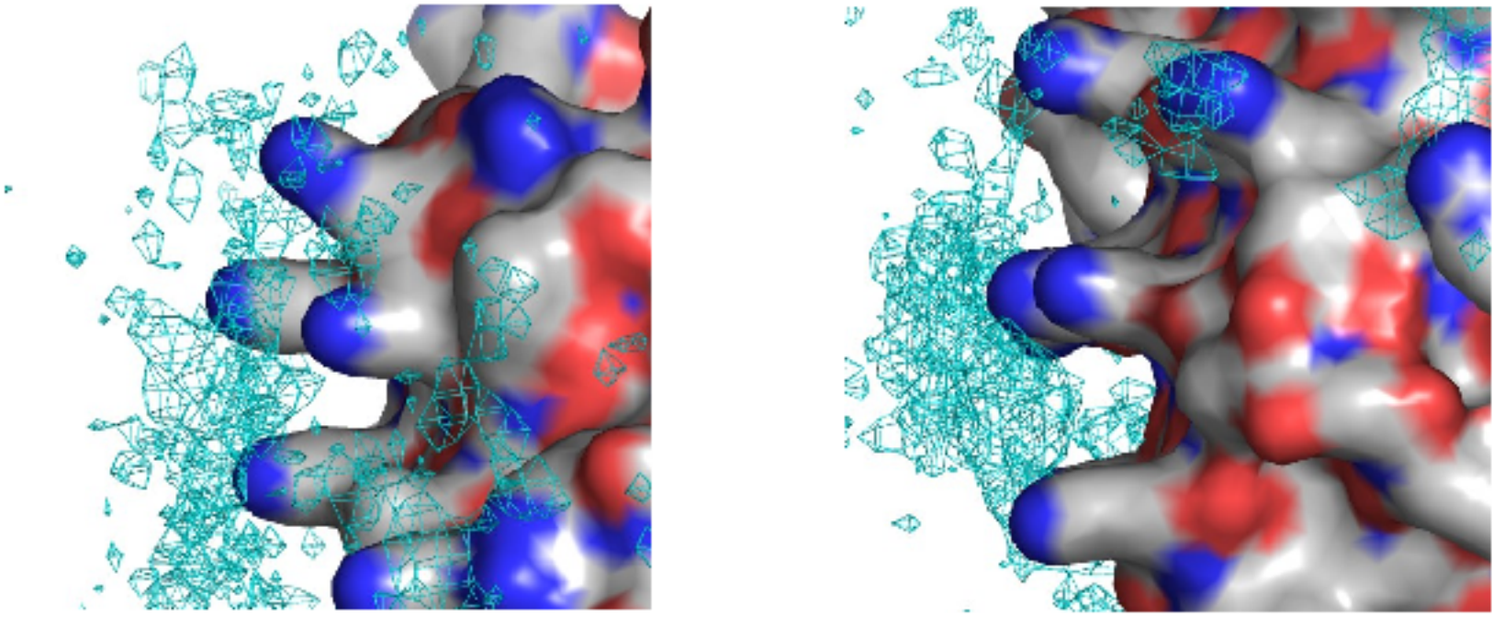
Figure 8



nGFP

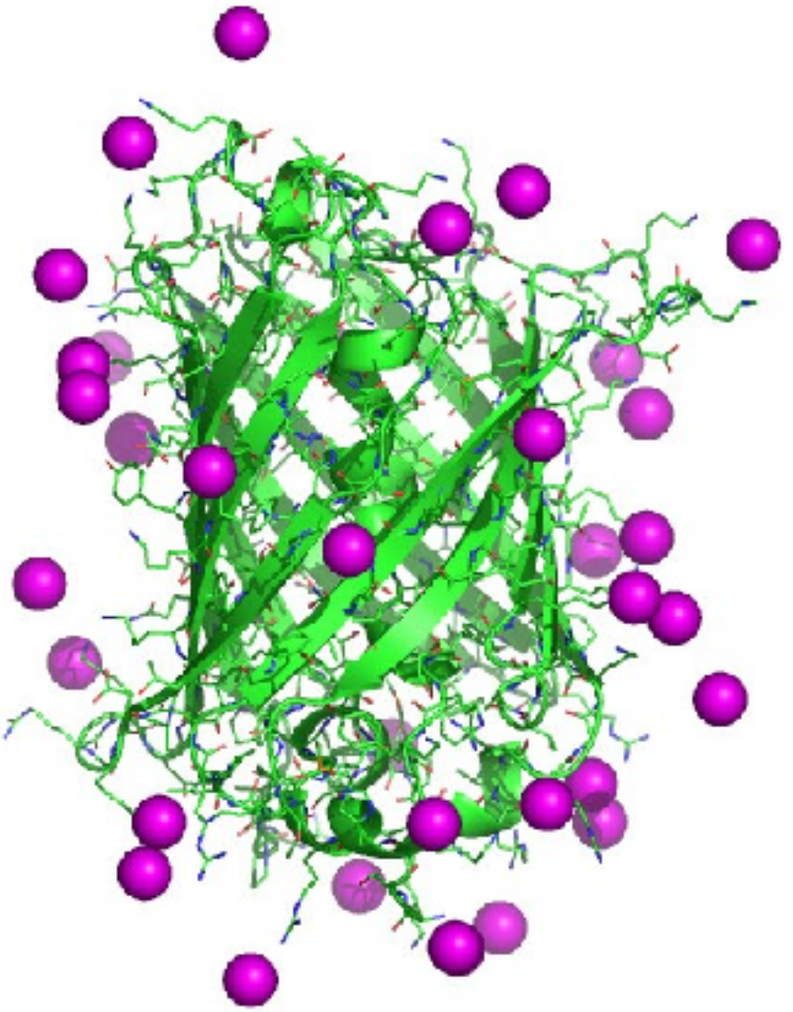
pGFP

Figure 9



Supplementary Figure 1

Supplementary Figure 2



Supplementary Figure 3